JOURNAL OF COMPLEXITY 8, 72–78 (1992)

Barnsley's Scheme for the Fractal Encoding of Images

S. GRAF

Fakultät für Mathematik und Informatik, Universität Passau, D-8390 Passau, Germany

Received April 19, 1991

1. INTRODUCTION

One major problem in image processing is the handling of the huge amount of data needed to record, store, or transmit a given (digital) image. Therefore one theme in image processing is to look for methods to reduce the amount of data needed to record an image up to a given accuracy, i.e., to look for methods of data compression. During the last decade Barnsley and his co-workers developed a scheme for data compression using fractal geometry. Here I outline the basic idea behind Barnsley's method. None of the results in this paper is new (see, for instance, Barnsley, 1988; Barnsley and Sloan, 1988). The paper is rather meant to draw the attention of people working on information-based complexity to this promising application of fractal geometry.

2. DETERMINISTIC ITERATED FUNCTION SYSTEMS

The basic mathematical notion in Barnsley's approach to data compression of images is that of an iterated function system. An *iterated function system* (IFS) is an N-tuple $(w_1, \ldots, w_N)$ of maps from a metric space (X, d) into itself. A map $w: X \to X$ is called a *contraction* iff there exists a constant $c \in \mathbf{R}$ with $0 \le c < 1$ and

$$d(w(x), w(y)) \le cd(x, y)$$

for all $x, y \in X$. The smallest c with this property is called the *Lipschitz constant* of w and is denoted by $\mathrm{Lip}(w)$. An iterated function system consisting of contractions $w_1, \ldots, w_N$ is called *hyperbolic*. In the present

0885-064X/92 $3.00

paper I consider only hyperbolic iterated function systems and, therefore, often omit the attribute "hyperbolic."

By $\mathcal{K}(X)$ we denote the set of all nonempty compact subsets of X. For $x \in X$ and $K \in \mathcal{K}(X)$ let

$$d(x, K) = \min\{d(x, y) : y \in K\}$$

be the distance from x to K. If $K, L \in \mathcal{K}(X)$ are given then

$$h(K, L) = \max(\max\{d(x, K) : x \in L\}, \max\{d(y, L) : y \in K\})$$

is called the *Hausdorff distance* of K and L. h defines a metric on $\mathcal{K}(X)$ and the metric space $(\mathcal{K}(X), h)$ is complete if and only if (X, d) is complete.

The following result is the cornerstone in the theory of iterated function systems.

2.1. *Existence of Attractors (Hutchinson, 1981)*

Let $(w_1, \ldots, w_N)$ be a (hyperbolic) iterated function system on a complete metric space (X, d). Then $W: \mathcal{K}(X) \to \mathcal{K}(X)$ defined by

$$W(K) = w_1(K) \cup \cdots \cup w_N(K)$$

is a contraction with respect to h with $\mathrm{Lip}(W) \leq \max(\mathrm{Lip}(w_1), \ldots, \mathrm{Lip}(w_n))$ and has a unique fixed point A in $\mathcal{K}(X)$.

The fixed point A of W is called the *attractor* of the IFS $(w_1, \ldots, w_N)$. The name originates from the fact that due to Banach's fixed point theorem, for an arbitrary $K \in \mathcal{K}(X)$, the iterates $W^n(K)$ converge to A.

For the considerations below we also need the following description of the attractor:

2.2. *A Cantorlike Construction of the Attractor*

If $X = \mathbf{R}^m$ and $(w_1, \ldots, w_N)$ is an IFS on X then there exists a compact nonempty subset J of $\mathbf{R}^m$ with $\mathrm{cl}(\mathrm{int}(J)) = J$ and $w_i(J) \subset J$ for $i = 1, \ldots, N$. For $(i_1, \ldots, i_n) \in \{1, \ldots, N\}^n$ set

$$J_{i_1, \ldots, i_n} = w_{i_1} \circ \cdots \circ w_{i_n}(J)$$

and

$$A_n = \bigcup_{(i_1, \ldots, i_n) \in \{1, \ldots, N\}^n} J_{i_1, \ldots, i_n}.$$

Then

$$A = \bigcap_{n \in \mathbf{N}} A_n.$$

As a consequence there exists a unique continuous map π: $\{1, \ldots, N\}^{\mathbf{N}} \to X$ with $\{\pi(\eta)\} = \bigcap_{n \in \mathbf{N}} J_{\eta_1, \ldots, \eta_n}$. Here $\{1, \ldots, N\}^{\mathbf{N}}$ carries the product topology (where $\{1, \ldots, N\}$ has the discrete topology). An alternative useful way to define π even if X is not equal to $\mathbf{R}^m$ is the following: Let $x \in X$ be arbitrary. Then for every $\eta \in \{1, \ldots, N\}^N$ the limit

$$\lim_{n \to \infty} w_{\eta_1} \circ \cdots \circ w_{\eta_n}(x)$$

exists and equals $\pi(\eta)$. For a detailed account of the results mentioned above as well as for standard examples the reader is referred to Barnsley's (1988) book.

3. The Encoding of Binary (= Black and White) Images

A binary image can be represented by a compact subset L of $\mathbf{R}^2$. Given such an image L and a $\delta > 0$ the problem is to find "nice" contractions $w_1, \ldots, w_N$ of $\mathbf{R}^2$ such that the attractor A of the IFS $(w_1, \ldots, w_N)$ satisfies $h(A, L) < \delta$. In this paper "nice" contractions are always affine contractions. An affine map from $\mathbf{R}^2$ into itself can be described by a matrix

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

and a translation vector

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix},$$

i.e., by six real numbers. Thus, if the attractor A of the IFS $(w_1, \ldots, w_N)$ satisfies $h(A, L) < \delta$ then the image L can be encoded by $6N$ real numbers up to accuracy δ. For every $\delta > 0$ such an IFS can be found but N may be very large. A problem related to complexity is to find upper and lower bounds for the smallest N with the above property and to construct applicable algorithms to obtain an IFS $(w_1, \ldots, w_N)$ whose attractor approximates the given binary image up to accuracy $\delta > 0$.

A heuristic tool for finding a coding of a given binary image L is provided by the following theorem

3.1. *Collage Theorem (Barnsley, 1988)*

Let $(w_1, \ldots, w_N)$ be a (hyperbolic) IFS on a complete metric space (X, d) with attractor A. Let $L \subset X$ be compact, $s = \max\{\mathrm{Lip}(w_1), \ldots, \mathrm{Lip}(w_N)\}$, and $W: \mathcal{H}(X) \to \mathcal{H}(X)$ be defined by $W(K) = w_1(K) \cup \cdots \cup w_N(K)$. If $h(W(L), L) < \varepsilon$ then

$$h(A, L) < \frac{\varepsilon}{1 - s}.$$

This result can be used in the following way. Given a binary image L one draws a rough outline of L and covers it by a number of smaller affine copies of itself. For each one of the smaller copies needed in the covering there is a (unique) affine map from the rough outline onto it. The affine maps determined by this procedure form the IFS one is looking for. (For a more detailed description of this method see Barnsley (1988).)

4. Random Iterated Function Systems and the Encoding of Gray Scale and Color Images

An IFS can only encode a black and white image. It is the purpose of this section to generalize the notion of an IFS in such a way that it allows the encoding of gray scale and color images.

A *random iterated function system* on a metric space X is an IFS $(w_1, \ldots, w_N)$ together with a probability vector $(p_1, \ldots, p_N)$. Each random IFS $(w_1, \ldots, w_N; p_1, \ldots, p_N)$ induces a Markov operator M on the space $P(X)$ of all Borel probability measures on X by the following identity:

$$M(\mu)(B) = \sum_{i=1}^{N} p_i \mu(w_i^{-1}(B)).$$

(Here B runs through all Borel subsets of X).

If X is compact, if $P(X)$ is endowed with the Hutchinson metric

$$d_{\mathrm{H}}(\mu, \nu) = \sup\left\{\int_X f\, d\mu \mid f\colon X \to \mathbf{R} \text{ s.t. } \forall x, y \in X |f(x) - f(y)| \le d(x, y)\right\},$$

and if $(w_1, \ldots, w_N)$ is hyperbolic then M is a contraction of the complete metric space $(P(X), d_{\mathrm{H}})$ (see Hutchinson, 1981).

The unique fixed point μ of M is called the invariant measure of the random IFS $(w_1, \ldots, w_N; p_1, \ldots, p_N)$. If $X \subset \mathbf{R}^m$ then this measure can also be obtained in the following way: Let π be the map from $\{1, \ldots, N\}^{\mathbf{N}}$ onto the attractor A of $(w_1, \ldots, w_N)$ as constructed in 2.2. Let $\tilde{\mu}$ be the product probability on $\{1, \ldots, N\}^{\mathbf{N}}$ obtained from the probability $(p_1, \ldots, p_N)$ on $\{1, \ldots, N\}$. Then μ is the image of $\tilde{\mu}$ with respect to π. If one wants the measure μ equally spread out on the attractor A (i.e., if μ is an "equidistribution" on A), one should choose the probability p_i approximately equal to the quotient "area of $w_i(A)$" divided by "area of A." Of course this works only if A has positive area. A gray scale image can be represented by a Borel probability on a compact subset of $\mathbf{R}^2$. Color images can be represented by three such probabilities each one standing for another primary color. A binary image is represented by an "equidistribution" on a compact subset of $\mathbf{R}^2$.

Given a gray scale image, i.e., a compactly supported probability ν on $\mathbf{R}^2$, the problem is to find a random IFS $(w_1, \ldots, w_N; p_1, \ldots, p_N)$ such that the corresponding invariant measure μ is close to ν with respect to the Hutchinson metric. As in the case of binary images the question arises of how big the random IFS must be to approximate μ up to a given accuracy. As in the case of sets one can prove a collage theorem for measures (see Barnsley, 1988, p. 365).

5. Reproduction of Images

Let μ be a Borel probability on $\mathbf{R}^2$ with compact support. We think of μ as representing an image. If we want to obtain a digitized version of μ we consider the following model: The screen of our computer is represented by a rectangle in the plane. This rectangle is divided into an array of smaller rectangles, the pixels. Each pixel B_{ij} is assigned the gray value $\mu(B_{ij})$. If our image is coded by a random IFS $(w_1, \ldots, w_N; p_1, \ldots, p_N)$ and we want to reproduce the digitized version of the image we have to find an algorithm to compute the values $\mu(B_{ij})$ approximately from the parameters of $(w_1, \ldots, w_N; p_1, \ldots, p_N)$. One possibile way to do this is provided by the following theorem.

5.1. *Theorem (Elton, 1987)*

If $(w_1, \ldots, w_N; p_1, \ldots, p_N)$ is a random IFS on a compact subset X of $\mathbf{R}^m$, $\tilde{\mu}$ the product measure on $\{1, \ldots, N\}^{\mathbf{N}}$ (see Section 4), μ the corresponding invariant measure, and B a Borel set in X, then for $\tilde{\mu}$ a.e. $\eta \in \{1, \ldots, N\}^{\mathbf{N}}$ and every $x \in X$,

$$\lim_{n\to\infty} \frac{1}{n} \operatorname{card}\{j \in \{1, \ldots, n\} | S_{\eta_j} \circ \cdots \circ S_{\eta_1}(x) \in B\} = \mu(B).$$

To reproduce a digitized image from its representation as a random IFS $(w_1, \ldots, w_N; p_1, \ldots, p_N)$ one, therefore, can proceed as follows (see Barnsley, 1988):

Let A be the attractor of $(w_1, \ldots, w_N)$. Choose a point $x_0 \in \mathbf{R}^2$ close to A (for instance, a fixed point of one of the w_i which automatically lies in A). Then choose a random index i_1 from $\{1, \ldots, N\}$ according to the distribution $(p_1, \ldots, p_N)$. Let $x_1 = w_{i_1}(x_0)$. Then choose a random index i_2 from $\{1, \ldots, N\}$ with respect to $(p_1, \ldots, p_N)$ independently of i_1. Let $x_2 = w_{i_2}(x_1)$. Continue this process until you have obtained a large number n of points.

If you want to reproduce a binary image color each pixel B_{ij} black for which there is an $x_\nu \in B_{ij}$. If you want to reconstruct a gray scale picture count the number of points from $\{x_1, \ldots, x_n\}$ which lie in B_{ij} and assign a corresponding gray value to B_{ij}.

The method for image reproduction described above is not very well suited for parallel implementation. A different algorithm connected with the theory of neural networks and suited for parallel implementation can be found in Bressloff and Stark (1991).

6. Comments

(a) The method described here should be illustrated by pictures. To avoid costly reproductions we refer the reader to Barnsley's (1988) book which contains numerous interesting pictures obtained from IFS representations.

(b) Of course the scheme for image encoding described above yields only a rough idea of what has to be done to encode "real" images by iterated function systems. To make the method applicable for practical purposes it has to be worked out to quite an extent. In particular the notion of an iterated function system has be generalized. Such generalized notions are, for instance, the place-dependent iterated function systems of Barnsley *et al.* (1988), the graph-directed constructions of Mauldin and Williams (1988) (see also Edgar, 1990), the sofic systems of Bandt (1989), and the recurrent sets of Dekking (1982) and Barnsley *et al.* (1989).

(c) A good survey on fractal geometry is Falconer's (1990) book.

References

Bandt, C. (1989), Self-similar sets. 3. Constructions with sofic systems, *Monatsh. Math.* **108**, 89–102.

Barnsley, M. F. (1988), "Fractals Everywhere," Academic Press, San Diego.

Barnsley, M. F., Demko, S. G., Elton, J. H. and Geronimo, J. S. (1988), Invariant measures for Markov processes arising from iterated function systems with place-dependent probabilities, *Ann. Inst. H. Poincaré* **24,** 367–394.

Barnsley, M. F., Elton, J. H., and Hardin, D. P. (1989), Recurrent iterated function systems, *Constr. Approx.* **5,** 3–31.

Barnsley, M. F., and Sloan, A. D. (1988), A better way to compress images, *Byte Mag.*, 215–223.

Bressloff, P. C., and Stark, J. (1991), Neural networks, learning automata and iterated function systems, *in* "Fractals and Chaos" (A. J. Grilly, R. A. Earnshaw, and H. Jones, Eds.), Springer, New York.

Dekking, F. M. (1982), Recurrent stets, Adv. Math. **44,** 78–104.

Edgar, G. A. (1990), "Measure, Topology and Fractal Geometry," Springer, New York.

Elton, J. (1987), An ergodic theorem for iterated maps, *J. Ergodic Theory and Dynamical Systems* **7,** 481–488.

Falconer, K. (1990), "Fractal Geometry," Wiley, Chicester.

Hutchinson, J. (1981), Fractals and self-similarity, *Indiana Univ. J.* **30,** 713–747.

Mauldin, R. D., and Williams, S. C. (1988), Hausdorff dimension in graph directed constructions, *Trans. Amer. Math. Soc.* **309,** 811–829.